

①

ITTC FILE COPY

AD-A198 937

Research Product 88-16

VARWARS Exercise Evaluation
Procedure Process
Rating Scales Users' Guide

ARI Field Unit at Fort Leavenworth, Kansas
Training Research Laboratory

July 1988

DTIC
ELECTE
SEP 21 1988
S E D



U. S. Army Research Institute for the Behavioral and Social Sciences

Approved for public release; distribution unlimited.

U. S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency under the Jurisdiction of the
Deputy Chief of Staff for Personnel

EDGAR M. JOHNSON
Technical Director

WM. DARRYL HENDERSON
COL, IN
Commanding

Technical review by

Mel Bohannon
Rex Michel



Accession For	
NTIS GRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution/	
Availability Codes	
Avail and/or	
Dist	Special
A-1	

NOTICES

FINAL DISPOSITION: This Research Product may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: This Research Product is not to be construed as an official Department of the Army document in its present form.

A10393

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified		1b. RESTRICTIVE MARKINGS	
2a. SECURITY CLASSIFICATION AUTHORITY		3. DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release; distribution unlimited.	
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE			
4. PERFORMING ORGANIZATION REPORT NUMBER(S)		5. MONITORING ORGANIZATION REPORT NUMBER(S) ARI Research Product 88-16	
5a. NAME OF PERFORMING ORGANIZATION U.S. Army Research Institute for the Behavioral and Social Sciences	6b. OFFICE SYMBOL (If applicable) PERI-SL	7a. NAME OF MONITORING ORGANIZATION --	
6c. ADDRESS (City, State, and ZIP Code) ARI Field Unit - Leavenworth P.O. Box 3407 Ft. Leavenworth, KS 66027-0347		7b. ADDRESS (City, State, and ZIP Code) --	
8a. NAME OF FUNDING / SPONSORING ORGANIZATION U.S. Army Research Institute	8b. OFFICE SYMBOL (If applicable) PERI-SZ	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER --	
8c. ADDRESS (City, State, and ZIP Code) 5001 Eisenhower Avenue Alexandria, VA 22333-5600		10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. P63744A	PROJECT NO. 2Q263 744A795
		TASK NO. 133	WORK UNIT ACCESSION NO. H0
11. TITLE (Include Security Classification) VARWARS Exercise Evaluation Procedure Process Rating Scales Users' Guide			
12. PERSONAL AUTHOR(S) Delane K. Garlinger and James W. Lussier			
13a. TYPE OF REPORT Final	13b. TIME COVERED FROM 03/87 TO 03/88	14. DATE OF REPORT (Year, Month, Day) 1988, July	15. PAGE COUNT 54
16. SUPPLEMENTARY NOTATION Also see: VARWARS: A Group Problem Solving Exercise, by James W. Lussier.			
17. COSATI CODES FIELD GROUP SUB-GROUP		18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number) Problem solving Evaluation Problem solving process Group problem solving	
19. ABSTRACT (Continue on reverse if necessary and identify by block number) The Director of the Combined Arms and Services Staff School (CAS ³) requested assistance in conducting an evaluation of the school's present curriculum for teaching group problem-solving skills. The request necessitated the development of a methodology to measure these skills. This report and its companion entitled, "VARWARS: A Group Problem Solving Exercise," document the methodology and provide guidance for implementation. This report contains rating scales tailored to the "VARWARS" problem. The scales are used to evaluate the actions taken by a group in developing a solution and certain behavioral descriptions to assist observers in making ratings. The manual also contains recommendations for use and suggestions concerning data analysis. Although the scales described in this manual are tailored to the VARWARS problem, most of the information can be used for evaluating other group problem-solving exercises. Footnotes indicate the changes necessary to adapt the scales to other problems.			
20. DISTRIBUTION / AVAILABILITY OF ABSTRACT <input type="checkbox"/> UNCLASSIFIED/UNLIMITED <input checked="" type="checkbox"/> SAME AS RPT. <input type="checkbox"/> DTIC USERS		21. ABSTRACT SECURITY CLASSIFICATION Unclassified	
22a. NAME OF RESPONSIBLE INDIVIDUAL Delane K. Garlinger		22b. TELEPHONE (Include Area Code) (913) 684-4933	22c. OFFICE SYMBOL PERI-SL

Research Product 88-16

**VARWARS Exercise Evaluation
Procedure Process
Rating Scales Users' Guide**

Delane K. Garlinger and James W. Lussier

ARI Field Unit at Fort Leavenworth, Kansas
Seward Smith, Chief

**Systems Research Laboratory
Robin L. Keesee, Director**

U.S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES

5001 Eisenhower Avenue, Alexandria, Virginia 22333-5600

Office, Deputy Chief of Staff for Personnel
Department of the Army

July 1988

Army Project Number
2Q263744A795

Training and Simulation


Approved for public release; distribution unlimited.

FOREWORD

The Fort Leavenworth Field Unit of the U.S. Army Research Institute for the Behavioral and Social Sciences supports the Combined Arms Center and its component activities with research on improving human performance in command and control (C²). Research areas include measurement of C² performance, enhancing human performance with computer-based decision tools, and procedures for command group training. Support is provided through long-term research projects and short-term Technical Advisory Service. This report is one of several prepared for the Combined Arms and Services Staff School (CAS³), a division of the Command and General Staff College (CGSC), which provides 9 weeks of training in staff skills to all U.S. Army Captains.

The Director of CAS³ requested assistance in conducting an evaluation of the effectiveness of the school's present curriculum for teaching problem-solving skills. This request necessitated the development of a methodology to measure these skills. This report, and its companion, entitled "VARWARS: A Group Problem Solving Exercise," document the methodology and provide guidance for implementation.

This project was funded under Research Task 1.3.3 (Improved Methods for Command Group Training) and the results were briefed to the sponsor in September 1987.


EDGAR M. JOHNSON
Technical Director

VARWARS EXERCISE EVALUATION PROCEDURE PROCESS RATING SCALES USERS' GUIDE

CONTENTS

	Page
INTRODUCTION	1
CHAPTER 1: GUIDELINES FOR OBSERVATION AND RATING	2
Recommendations:	
#1 Become familiar with rating scales	2
#2 Be informed about rater error and biases	2
#3 Raters should be unaware of status or condition of group being rated	3
#4 Keep raters consistent for groups to be compared	3
#5 Use Modified Delphi Technique to establish scores	4
#6 Videotape group problem solving process	4
CHAPTER 2: RATING SCALE FORMAT AND SCORING	6
Format	6
Scoring	6
CHAPTER 3: UNDERSTANDING THE RATING SCALE	8
Getting Organized	8
Information Sharing and Coordination	13
Decision Making	16
Professionalism	20
Leadership	24
CHAPTER 4: DATA ANALYSIS	30
Suggestions:	
#1 Use aggregate scores	30
#2 Use multivariate procedures	31
#3 Assess rater reliability	31
REFERENCES	32
APPENDIX A. RATER TRAINING MATERIAL	A-1
B. PROCESS RATING SCALE	B-1
C. INTERRATER RELIABILITY PROGRAM	C-1

Introduction

The VARWARS Problem is a group planning and resource allocation exercise developed by the Army Research Institute at the request of the Director of the Combined Arms and Services Staff School (CAS³). It meets CAS³'s requirement for a quantitative, objectively scorable exercise which may be used for:

- o formal evaluation of instructional effectiveness
- o training of teamwork in problem solving
- o tailoring teaching to deficient problem solving skills

This manual contains rating scales tailored to the VARWARS Problem. They are used for evaluating the actions taken by a group in developing a solution. A companion volume "VARWARS: A Group Problem Solving Exercise" tells how to administer the problem, how to score it, and how to interpret the objective score (Lussier in press). In contrast, this volume says what you should look for in problem solving groups, whether your purpose is evaluation or diagnosis. It also provides rating scales to help structure your observations. Although the scales are tailored to VARWARS, most of the information can be used for evaluating other group problem solving exercises. Footnotes indicate which changes would be necessary to adapt the scales to other problems.

Chapters 2 and 3 cover the essential information for using the process scales in training. Chapters 1 and 4 include additional information necessary for using the scales in a formal program evaluation. Appendix B contains the forms to be used in rating group processes.

Chapter 1: Guidelines for Observation and Rating

Recommendations:

Although the rating scales are, for the most part, defined in behavioral terms, there is still some degree of subjective judgment required by raters. For this reason, it is suggested that the recommendations presented below be followed in order to maximize the reliability and accuracy of the ratings. Suggestion #'s 1 and 2 are recommended for any application of the rating scales. However, suggestion #'s 3, 4, 5, and 6 are most applicable when the rating scales are to be used for formal evaluation purposes.

1

**Become familiar
with rating scales.**

It is recommended that personnel who serve as raters be knowledgeable in regards to the processes to be rated. An explanation of the rating scales is presented in Chapter 3 of this manual. Raters should review and become familiar with the scales before attempting to use them.

2

**Be informed about
rater error and biases.**

Even experienced raters often make systematic rating errors which result from misperceptions and individual tendencies. Evaluators are usually unaware that they are making these judgment errors, but once they become familiar with the common causes of error they can usually correct themselves and achieve

a functional level of accuracy in their rating performance.

Therefore, everyone who will serve as a rater should be informed about common rater errors and biases. Appendix A contains descriptions and examples of the most common types of bias which can adversely affect the rating process (Garlinger, in press). It is recommended that this material be reviewed by individuals who participate as raters. Ideally, the material should be reviewed as near to the beginning of the rating session as possible. Also, individuals who will serve as raters for a number of sessions over a period of time should review the material periodically as a reminder of the potential hazards.

3

**Raters should be
unaware of status
or condition of
group being rated.**

It is strongly recommended that the individuals who serve as raters be blind to the condition or status of the group being rated. For instance, if the rating scales are being used in an assessment of instructional effectiveness, raters should be unaware of which groups have had instruction and which groups have not.

4

**Keep raters
consistent for
groups to be
compared.**

When comparisons are to be made between groups, the same raters should assess all groups. By keeping raters consistent, any individual bias in ratings will probably be uniform across conditions, thereby reducing the possibility of spurious results.

5

Use modified Delphi technique to establish scores.

In order to insure that ratings are as reliable as possible, a modified Delphi technique is advocated to establish the score on each process item. The technique requires that multiple raters observe and make their individual ratings on each item. After the initial rating, the observers meet and report their individual ratings on each item. If there is not complete agreement on the rating for an item, raters report their reason for the rating given. After all raters have given their rationale for the rating of that item, each rater again rates the item in view of the information and rationale provided by his co-raters. In order to prevent the process from becoming unduly time consuming, if consensus is not reached on the second rating iteration the average of the ratings should be awarded for that item.

The interrater reliability of the initial ratings can be examined by computing an intraclass correlation coefficient (Kerlinger, 1974). However, the process can be simplified by use of the Reliability procedure of SPXX-X (SPSS, 1983). The organization of the data and the program for running this procedure is provided in Appendix C.

6

Videotape group problem solving process.

Although it is possible for raters to make their initial rating while observing the group problem solving process directly, it is recommended that the process be videotaped so that the raters

have the opportunity to review those aspects about which there may be confusion or disagreement. It would also be possible to make initial ratings directly from the videotape rather than from the live performance. Videotaping allows greater control of order and sequence effects in rating and simplifies the use of a blind rating procedure.

Chapter 2: Rating Scale Format and Scoring

Format

The rating scales measure the group problem solving process in regards to five general scales: getting organized, information sharing and coordination, decision making, professionalism, and leadership. Each of these five scales is composed of items which measure four specific dimensions which the authors considered to be most relevant to each scale. The complete rating form is included in this report as Appendix B.

Each scale and dimension will be discussed in more detail in Chapter 3. The discussions will include an explanation of each scale and its dimensions, as well as a description of typical behavior associated with the numerical rating points of each dimension. Although the descriptions are, for the most part, self-explanatory, specific examples of behavior associated with the numerical rating points will be provided where necessary for clarity.

Scoring

The dimensions which are associated with the information sharing and coordination scale are scored through use of a checklist for twenty items of information which should be shared between members of the group. One point is given for each item checked; therefore, the maximum score for this scale would be twenty points.

Each of the remaining four scales are measured with four items each which represent the four dimensions of that scale. Each item is scored from one to five, based on behavioral descriptions which accompany each numerical rating. A rating of five indicates

superior performance on that dimension and a rating of one represents a very low level of performance. The maximum number of points for each of these scales is also 20 points. Therefore, the maximum number of points possible on the entire process rating form is 100.

Chapter 3: Understanding the Rating Scale

1. Getting Organized

This scale measures the thoroughness and efficiency of problem solving groups in executing the preliminaries which provide the foundation for the group problem solving activities to follow. The VARWARS problem is a very complex one, and the time limit for reaching and documenting a solution is only three hours. Therefore, it is important that groups organize quickly and successfully. The dimensions measured with this scale are: *organizing into teams and assigning roles, developing a time management plan, defining and restating the problem, and preliminary information gathering.* Each of these dimensions will be discussed below.

a. Organizing into Teams and Assigning Roles.

a. Organizing into Teams and Assigning Roles.

- (1) No roles or teams were established.
- (2) There was a division of tasks, but no clear assignment of individual responsibility.
- (3) Leader assigned roles arbitrarily without consideration of preferences or qualifications.
- (4) Assignments were based on preference alone OR the process of determining qualifications was excessively time consuming.
- (5) Leader assigned roles based on preferences and qualifications.

The VARWARS exercise provides instructions and unique information for three separate teams and for the group leader. The three teams are called the training team, personnel team, and budget team, and each is provided with information unique to their task. The exercise was developed for a group size of seven - the leader plus three teams composed of two members each. (Instructions can be modified to accommodate larger or smaller groups). During the "getting organized" phase, individuals should be assigned to one of the three teams based upon some reasonable rationale, preferably experience in the area of training, budget or personnel. The group leader is designated prior to the start of the exercise based upon rank or date of rank - the senior member of the group is always the group leader regardless of previous experience in group leadership. The rating points are mostly self-explanatory. Obviously, the worst situation would be a group which did not attempt to make use of the team division of responsibility and worked in undifferentiated chaos. Ideally, the leader would quickly ascertain what previous experience or education each member may have which would provide expertise in either training, budget, or personnel, and then assign individuals to teams accordingly. For individuals with no experience in any of these areas, the assignment could be made based on the preference of that individual. However, the process of organizing the teams should not become overly time consuming. For instance, some leaders may have all members provide a lengthy description of him/herself which goes beyond the requirements of the problem. This is an inefficient process for an exercise with a three hour time constraint.¹

¹ Since the process rating scales were designed for use with ad hoc groups, some modifications would be required in order to use this item for an established group in which the leader was already knowledgeable about the members' expertise and preferences. In such an instance it may not be possible to ascertain the criteria used by the leader in making team assignments.

**b. Time
Management
Plan**

b. Time Management Plan.

- (1) No attempt made.
- (2) Overall time limit mentioned only.
- (3) Discussed need for time plan, various options considered, no decision reached and clearly stated.
- (4) Developed time plan that was not fully adequate OR process was inefficient and overly time consuming.
- (5) Arrived at a clearly-defined workable plan or schedule that was clearly understood by all members.

In order for the group problem solving process to proceed efficiently, the team should establish a plan or schedule to insure timely completion of component activities. The three hour time constraint imposed for reaching and documenting a decision makes this aspect of getting organized crucial to success.

The lowest rating would be given to a group which did not recognize the need for or attempt to formulate any sort of schedule for their progress. Although mentioning of the overall time limit is of little benefit in regards to insuring that the process proceeds in a timely manner, it is minimally superior to not addressing the issue at all. A mid-scale rating would be given to a group that recognized the need for a time management plan or schedule, but which did not get beyond discussing the need and failed to establish such a plan. Ratings of four and five are primarily distinguished by the adequacy of the plan developed. If a group establishes a plan which includes only a portion of the processes involved in reaching a solution, or accounts for only a portion of the time available, then they would receive a rating of four. To receive a rating of five, the plan should include a schedule of

activities to take place up to and including completion of the forms which document their solution. The plan need not be elaborate and should not contain great detail, but must encompass activities from start to finish of the problem solving process.

**c. Problem
Definition and
Restatement**

c. Problem Definition and Restatement.

- (1) Made no attempt at problem definition.
- (2) Made reference to need only but did not accomplish problem definition.
- (3) Reached an unsatisfactory problem definition.
- (4) Problem definition was only adequate OR process was inefficient and time consuming.
- (5) Efficiently arrived at a good problem definition which was beneficial to problem solving process.

This item is concerned with how well the groups are able to integrate the preliminary information available to them and formulate it into a definition of the problem or goal statement. Establishing the definition of the problem is important because it should provide the focus and direction needed to achieve a successful solution.

Rating points one and two are self-explanatory. However, rating points three through five are more ambiguous as to what constitutes an unsatisfactory, adequate, or good problem definition. The following examples are provided for clarity:

Example:

Unsatisfactory: Select the device to provide training.

Adequate: Provide the most training possible in 150 weeks.

Good: Maximize score points by scheduling the most training for units according to priorities given, on a fixed budget, and in a three year time frame.

**d. Preliminary
Information
Gathering**

d. Preliminary Information Gathering.

- (1) No information from individual team instructions was presented or requested.
- (2) Information was requested but preliminary information gathering process was generally unproductive.
- (3) Information was provided from one or two team forms only.
- (4) Sufficient information was provided but process was unduly time consuming with too detailed a level of information provided.
- (5) All teams and leaders provided sufficient information to begin problem solving.

This item is concerned with assessing the adequacy of the group's preliminary sharing of information. At this point in the problem solving process, the getting organized stage, the emphasis should be on providing a general overview of each team's area of responsibility, and the types of unique information each has. This

phase of information gathering, which should be completed by 30 to 45 minutes after the start of the exercise, should not be confused with the information sharing which occurs later in the problem solving process and which will provide more specific and detailed information. During the getting organized stage, the group should be most concerned with disseminating enough information so that they may begin the problem solving process.

The descriptions associated with the rating points for this item are self explanatory. Maximum score is given to groups which provide sufficient information from the four sources (training, personnel, and budget teams and group leader) for the group to proceed, but which do not use an excessive amount of time to do so. Groups should be able to share an adequate amount of information for preliminary purposes in 10 to 15 minutes.

2. Information Sharing and Coordination

The VARWARS problem is complex and requires integration of many specific pieces of information. Each of the three teams (training, personnel, and budget), as well as the leader, receive a unique portion of that information. Over the course of the exercise it is necessary for all teams and the leader to share the unique information they possess and the estimates they develop with other group members. This scale contains a checklist of several items of specific information, the source of each piece of information, and who should receive that information². The timeliness of the information sharing and coordination is not measured by this scale in an absolute sense; however, in order to receive credit for passing an item of information, it must be received by the recipient prior to the selection of a course of action

² If these process rating scales are used with problems other than VARWARS, the information sharing checklist must be tailored to the problem being used.

(COA). Information received after the COA has been selected cannot contribute to problem solution.

Individuals who will serve as raters should carefully read the VARWARS problem instructions and information packages to become familiar with where information originates, where it should go, and why.

**Source and
Recipients of
Critical Elements
of Information**

Source and Recipient of Critical Elements of Information				
	L	B	T	P
Number of Devices		S	R	R
Number of Trainees and Sessions		S	R	
Point Cutoff Scores for Units	S		R	
Device Delivery Dates		S	R	R
Operator Availability Dates			R	S
Ratio of Weekday to Weekend Personnel		R	S	
Importance of Corps Training	S		R	
Device Maintenance Schedule		S	R	R
Relative Importance of Active and Reserve	S		R	
Number of Operators and TM per Session		S		R
Ratio of AC to RC Training Planned			S	R
Importance of AC and RC	S			R
Prototype Devices Available		S		R
Estimate of TDY Costs		R	S	
Estimate of Amount of Training Required		R	S	
Estimate of Personnel Costs		R		S
List of Units not Scheduled	R		S	
Key: L = Leader P = Personnel Team B = Budget Team S = Source T = Training Team R = Recipient				

3. Decision Making

This scale is intended to evaluate the core decision making activities, and includes items measuring four dimensions of decision making: *development of courses of action, generation of criteria, analysis of courses of action, and selection of a course of action*. These dimensions will be discussed below.

a. Development of Courses of Action

a. Development of Courses of Action.

- (1) Group did not arrive at any overall course of action but worked on completely unintegrated subsolutions.
- (2) Proposed only one course of action which was accepted with insufficient analysis.
- (3) Course of action generation proceeded sequentially. A single course of action was analyzed for feasibility before alternatives were considered.
- (4) Proposed two or more courses of action that were not sufficiently distinct or were inadequate in concept OR group was inefficient and used excessive time in generating courses of action.
- (5) Proposed two or more adequate and distinct COA.

Based upon the information available to them and their problem definition or goal statement, the group should work together to generate several potential courses of action for consideration.

The lowest score on this dimension would go to the group which never comes together to discuss alternatives, but instead each

team chooses alternatives from their unique viewpoint of the problem. Reaching workable alternatives requires an overall perspective which can only be provided if the whole group works together, with each team offering their perspective of constraints and options.

The most successful groups will consider a number of alternatives rather than hastily accepting a single alternative. Ideally, groups will generate at least two distinct alternatives which can be evaluated against the criteria to be established by the group in order to choose the best course of action.

b. Generation of Criteria for Selection

b. Generation of Criteria for Selection.

- (1) No criteria established. Group seemed unaware of need for criteria.
- (2) Discussed need for criteria, but none established.
- (3) Criteria were proposed but group failed to reach consensus or to achieve a common understanding of the criteria.
- (4) At least one relevant criterion was established but clearly relevant criteria were ignored OR process was inefficient and overly time consuming.
- (5) Criteria were relevant, clearly defined and no clearly relevant criteria were ignored.

In order to evaluate the desirability of the alternative courses of action that are generated, the group must develop criteria with which to measure and compare the alternatives.

Rating points one through three are self-explanatory, and reflect levels of failure in regards to criteria development. Rating points

four and five, however, are less clear. Both of these rating points require subjective judgment by raters in determining whether all clearly relevant criteria are established. The relevant criteria will, of course, be related to the problem definition or goal statement. Therefore, it is not possible to supply an exhaustive list of clearly relevant criteria which would apply to every group. However, cost and number of people trained are generally criteria which are relevant to the solution of the VARWARS problem, regardless of the group's formulation of the problem. Raters must decide if there are additional criteria which are clearly relevant to the group's unique problem definition or goal statement.

**c. Analyze Courses
of Action and Prepare
Estimates**

c. Analyze Courses of Action/Prepare Estimates.

- (1) No estimates made OR estimates were based on meager analysis.
- (2) All teams began work but none completed a satisfactory analysis.
- (3) Only one team analyzed courses of action sufficiently.
- (4) Two teams analyzed courses of action sufficiently.
- (5) Each course of action was analyzed completely by all teams. Costs and benefits were accurately identified.

Once the alternative courses of action have been selected and the criteria identified, the group should return to their individual teams to prepare estimates on the alternatives. For instance, the budget team should prepare an estimate of the purchase costs associated with each of the alternatives; the personnel team should prepare estimates of the number and costs of the training personnel required to support each of the alternatives, etc.

Estimates should be based upon actual calculations rather than "hunches," although simplifying assumptions could be used to develop preliminary estimates quickly.

The rating scale reflects varying degrees of success in regards to this dimension - from not making estimates to complete estimates by all teams. Rating is rather straight-forward and involves the observation of the individual teams in arriving at their estimates.

d. Selection of Courses of Action

d. Selection of Courses of Action.

- (1) No decision reached.**
- (2) Arbitrary decision reached without consideration of pros and cons.**
- (3) Discussed pros and cons of the alternatives, but reached a decision without use of a comparative method.**
- (4) Used comparative method without establishing weights of criteria.**
- (5) Used comparative method to select a course of action based on an analysis of established, weighted criteria.**

Once the alternative courses of action have been generated, the criteria for selection identified, and the team analyses of the alternatives are completed, the group must put all these together to select the most desirable course of action.

This dimension evaluates the extent to which the group uses a comparative method to make a selection of one course of action. Ideally, the criteria will be weighted in accordance with importance to problem solution.

4. Professionalism

This scale measures the extent to which the group displays conduct which conforms to high standards of dedication and effort. The scale includes the dimensions of: *dedication to task, error checking, incremental improvement of complete solution, and the ability of the group to work within deadlines and manage time.* Each of these dimensions will be discussed below.

a. Dedication to Task

a. Dedication to Task.

- (1) Excessive amounts of irrelevant behaviors or idleness which seriously impaired group performance.
- (2) Moderate amounts of irrelevant behavior or idleness which had some negative impact on performance.
- (3) Some members failed to maintain attention to task despite prodding.
- (4) All team members worked continuously but some required prodding by others.
- (5) Entire group showed enthusiasm and attention to task throughout.

This dimension is concerned with the degree to which group members attend to the activities relevant to problem solution rather than engaging in irrelevant behavior and conversation.

The raters must decide how much irrelevant behavior is required to warrant a rating of "excessive amounts" rather than "moderate amounts." In general, if, in the judgment of the raters, the problem solving process is handicapped by the amount of irrelevant behavior, then the "excessive" rating should be given. Since the exercise takes three hours to complete, it can be expected that group members will leave the room once or twice for short breaks

(5 minutes or less). However, an excessive number of absences, or absences of excessive length of time, should be considered irrelevant behavior. Also, raters should attend to conversations taking place within the teams during those times when the teams are required to work individually. During such times, irrelevant conversations may hamper success on the team task. In addition, look for any individuals who sit idly while the rest of the group works, as idleness should be considered as irrelevant behavior.

b. Error Checking

b. Error Checking.

- (1) No discernable checking of facts, conclusions, methods, etc., occurred.
- (2) Too infrequent error checking clearly resulted in negative effect on solution.
- (3) Error checking was evident but the overall process was inadequate and at least one major error was uncorrected.
- (4) Error checking process proceeded satisfactorily with no major errors undiscovered, but process was inefficient and overly time consuming.
- (5) Frequent error checking led to positive effect on outcomes, without becoming unduly time consuming.

The VARWARS problem is very complex and uses a large amount of detailed information. Group members should be concerned with catching any errors or misunderstandings which may be a hazard to problem solution.

Ratings one through three are rather self-explanatory and are concerned with the degree of error checking ranging from none at all to moderate but inadequate. The difference between rating points four and five is the difference between being careful and being compulsive. Error checking is a beneficial practice in

problems such as VARWARS, but too much of a good thing can cripple performance. For example, recalculating a long computation to double-check accuracy is inefficient and overly time consuming.

**c. Incremental
Improvement
of Complete
Solution**

**c. Incremental Improvement of Complete
Solution.**

- (1) Group quit immediately after solution was reached.
- (2) Only 1 or 2 members continued to work after solution was reached.
- (3) Group continued to discuss final solution without making improvements.
- (4) Group continued working with minor improvements.
- (5) Group continued work after complete solution was reached with significant improvement.

When the group reaches a solution as a result of their decision making process, they should continue to "fine-tune" that solution as long as time permits.

All rating points are self-explanatory; however, raters will be required to decide what constitutes minor and significant improvement. The following examples are provided for clarification:

Examples:

Major Improvement: (1) Add a machine and additional training sessions when money left over. (2) Realign personnel budget to have fewer weekend personnel. (Change from 2x to 3/2x).

Minor Improvement: (1) Change priorities on training assignments to give closer units more sessions (stretches TDY budget further). (2) Fill small gaps in training schedule with unassigned units.

**d. Work Within
Deadlines and
Manage Time**

d. Work Within Deadlines and Manage Time.

- (1) Group failed to establish or meet deadlines or showed no awareness of deadlines.
- (2) Group showed awareness of deadlines but seldom succeeded in meeting deadlines.
- (3) Sometimes met deadlines.
- (4) Usually met deadlines.
- (5) Members performed tasks to meet all established deadlines.

This dimension measures the groups' adherence to established deadlines. Groups should be conscientious about conducting tasks as efficiently as possible in order to stay within time limits and maintain progress toward a solution.

The difficulty here is defining seldom, sometimes, and usually. The group which fails to establish deadlines should receive the lowest score on this dimension. For those groups which do establish deadlines, rating points two through five differentiate degree of success in meeting them. The suggested rule of thumb for distinguishing whether a rating of seldom, sometimes, or usually should be given, is that if a group meets their deadlines

less than 1/3 of the time, a rating of 2 (seldom) should be given; if deadlines are met from 1/3 to 2/3 of the time, a rating of 3 (sometimes) should be given, and a rating of 4 (usually) would be given to groups which were successful in meeting their deadlines more than 2/3 of the time but did not achieve 100% success.

5. Leadership

This scale is concerned with assessing the performance of the designated group leader in bringing together individuals to work effectively toward a common group goal. The concern here is not to evaluate the leader's style of leadership, but rather to assess his effectiveness in developing and maintaining a cohesive group effort. Although good leadership encompasses many dimensions, the four which have been chosen for inclusion in this category are: *providing focus and direction to group effort, making and stating decisions clearly, conflict resolution, and keeping the group to an appropriate time schedule.*

a. Provide Focus and Direction

a. Provide Focus and Direction.

- (1) Group was directionless. Never seemed to know what they should be doing or how to do it.
- (2) Leader made suggestions only or gave inadequate guidance. Group was essentially directionless.
- (3) Leader provided some focus. Group generally understood what they must accomplish, but not how to do it.
- (4) Leader provided focus and direction generally, but occasionally failed to redirect effort when group was confused or followed inappropriate or unproductive paths.

(5) Leader always maintained clear focus and direction of group. Quickly brought group on target.

With a problem as complex as the VARWARS problem, it is essential that the leader provide direction for the group effort and keep the group focused on the common goal. Since each team has a unique perspective of the problem it is easy for the group to fall into a situation where the teams are working in an unintegrated manner or get so caught up in the mechanics of their tasks that they lose sight of the "big picture." The leader must see that this does not happen.

The rating points reflect the degree to which the leader is successful in keeping the group oriented and working productively toward problem solution, and ranges from "never" to "always" provided focus and direction, with ratings between the two extremes reflecting increasing degrees of success. In distinguishing between the ratings of 2 and 3, the rater must decide whether the group has a general understanding of their task, which must be inferred from observing the directions provided by the leader and the group behavior which results. To receive a rating of 4, the leader must clearly have provided focus and direction, but failed to redirect teams or individuals who wandered off on irrelevant tangents on one or two occasions.

**b. Made and Stated
Decisions Clearly**

b. Made and Stated Decisions Clearly.

- (1) Appeared unable to make decisions.**
- (2) Decisions were not communicated to group or were made reluctantly.**
- (3) Attempted to communicate decisions, but left misunderstanding and confusion.**

- (4) Decisions were presented incompletely and required group effort and questioning to reach understanding.**
- (5) Decisions were made appropriately and clearly communicated to group.**

The problem solving process presents points at which decisions must be made in order to progress toward solution. The leader should make these decisions based upon information and analyses provided by the teams. A leader who does not recognize the need for a decision or who does not communicate the decision to the group can "derail" progress and much time can be lost in confusion.

Assessing the leader's performance on this dimension requires that raters recognize the points during the exercise where a decision is clearly required, and then observe the leader's behavior in response to the situation. If the leader consistently fails to recognize the need for a decision or is not able to provide a decision, then the lowest rating (1) is indicated. If the leader attempts to make decisions but is not completely successful because of poor communication skills or other reasons, then ratings of 2, 3, or 4 would be given, as appropriate. Only those leaders who consistently make and communicate clear and timely decisions should receive the top rating of 5.

c. Conflict Resolution

c. Conflict Resolution.

- (1) Leader made no attempt to resolve conflicts where leader intervention was clearly indicated.**
- (2) Attempts to resolve conflicts failed.**
- (3) Conflicts were resolved arbitrarily to detriment of group performance.**

- (4) Most conflicts were resolved but required excessive time.**
- (5) All conflicts resolved smoothly so that there were no detrimental effects on group performance.**

The number of conflicts which arise during the problem solving exercise will vary from one group to another. Some groups will have many conflicts whereas others will have few, or none. Whether the conflicts are few or many, however, the leader should recognize when a conflict exists and take appropriate steps to resolve the issue.

The key to assessing leader behavior on this dimension is identifying conflicts when they occur. The term conflict as used here refers to more than open hostility between group members. Other, more subtle behaviors should also be considered as conflicts, such as a group member who attempts to usurp the leadership position, and power-plays between group members. Also, if one or more group members fail to participate and sit idly while others work, this should be considered a conflict, even if the group members who are working appear not to notice or do not overtly confront the unproductive member.

The number of conflicts which arise should have no impact on the rating of the leader's performance on this dimension, as the number will depend more upon the characteristics of the individuals than the quality of the leadership. Rather, the leader should be rated on his ability to resolve conflicts when they arise. However, it is possible that some groups will have no incidents of conflict. In such a case, the average of the other three leadership dimensions should be given to this item. For instance, if a group displayed no conflict behavior and ratings of 4, 3, and 5 were given for the dimensions of providing focus, making decisions, and keeping to time schedule, respectively the rating of 4, the average of these three ratings, would be given for this dimension.

**d. Kept Group to
Appropriate
Time Schedule**

d. Kept Group To Appropriate Time Schedule.

- (1) Leader made no attempt to keep group to any schedule when group clearly required such guidance.**
- (2) Leader tried to maintain orderly progress but failed.**
- (3) Leader allowed group to miss deadlines and occasionally fall behind so that some tasks were hurried.**
- (4) Leader supervised progression through tasks adequately.**
- (5) Strong leadership kept group to timely and orderly sequence of tasks. Group had sufficient time for all phases.**

The VARWARS exercise time limit of three hours makes it essential that groups conscientiously pace themselves according to a time plan or schedule. For this dimension, a formal time plan is not required to receive a high rating. The leader may have his own mental time plan rather than one that is explicitly stated during the organization phase of the problem solving process. What is important is whether the leader tells the group when particular tasks must be completed and then enforces the deadlines. For instance, if the leader asks the teams to perform analyses and estimates, he should also state a deadline for completion and then supervise team efforts to insure compliance.

Rating points one through three are self-explanatory. However, rating points four and five require that the raters make a judgment about the group as well as the leader. Since successful leadership on this dimension requires more skill when leading a

sluggish, less motivated group than when leading a highly motivated group, rating points four and five are intended to make this distinction. If the group is highly motivated and leadership requirements are low, successful performance on this dimension should be given a rating of four. The rating of five is reserved for those individuals who are successful on this dimension as a result of strong leadership of a poorly motivated group.

Chapter 4: Data Analysis

Suggestions:

When the rating scales are used for classroom training, the scores are intended to assist instructors in conducting their After Action Review (AAR) and to guide discussion of the exercise. However, if the scales are used for a formal evaluation, a more elaborate data analysis may be desired. The choice of procedures used to analyze the process scores will largely depend upon the purpose(s) for which the evaluation was conducted. If the purpose of the evaluation was to assess the effectiveness of problem solving instruction, a multivariate between-groups analysis of variance would be the appropriate procedure. On the other hand, if the intent was to examine the relative contribution of the process categories to the overall problem solution score then a regression analysis would be conducted. It is not the intent of this manual to provide detailed instruction for conducting the various statistical procedures one may want to use for analysis of the process scores, but rather to provide some suggestions which may be helpful regardless of procedure used.

#1 Use Aggregate Scores

The first suggestion is that the aggregate process scale scores and the total of the process scores be used for data analysis rather than attempt to conduct analyses of individual items. This suggestion is made because the scale aggregates and total scores have more reliability than any individual items, and also this avoids conducting analyses of twenty or more dependent variables which would increase the possibility of finding spurious statistical significance.

2

Use Multivariate Procedures

The second suggestion is that multivariate analysis procedures be used whenever possible instead of a univariate approach. This suggestion is based upon the certainty that the five category scores will be highly correlated. Therefore, if each score is analyzed separately, the variance which the scores share will be analyzed redundantly, resulting in increased possibility of spurious significance. (Barker and Barker, 1984).

3

Assess Rater Reliability

The final suggestion is that the reliability of the raters' original judgments be assessed. Although the final scores are based upon the arrived-at consensus of the raters rather than an average of the original ratings, a reliability check of the original ratings will provide a gauge of the probable accuracy of the final ratings. If only two raters are used, their interrater reliability can be checked by simply conducting a Pearson correlation of their ratings on the individual items or aggregate scale scores. However, when three or more raters are used, as is recommended, then the procedure requires use of an intraclass correlation analysis. The formula for intraclass correlations can be found in most advanced statistics books. However, a simple method for analyzing the reliability of three or more raters is to use the Reliability Procedure of SPSS^X (1983) and code individual ratings as items. This is somewhat different from the normal use of the Reliability Procedure. Therefore, a programming outline of instructions for using the Reliability Procedure of SPSS^X in determining interrater reliability is contained in Appendix C.

References

- Barker, H. R., and Barker, B. M. (1984). Multivariate Analysis of Variance (MANOVA): A Practical Guide in Its Use and in Scientific Decision Making. University of Alabama Press.
- Garlinger, D. K. (1988). The effectiveness of a rater training booklet in increasing accuracy of performance ratings. (Research Note 88-27). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A194 723)
- Kerlinger, F. N. (1973). Foundations of behavioral research. New York, NY: Holt, Rinehart, and Winston, Inc.
- Lussier, J. W. (in preparation). VARWARS: A group problem solving exercise. (Research Product, in press). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- SPSS, Inc. (1983). SPSS^X user's guide. New York: McGraw-Hill Book Company.

APPENDIX A
RATER TRAINING MATERIAL

Obstacles to Rating Accuracy

When an individual observes and evaluates others, systematic rating errors often occur. Rating errors are mistakes in judgment resulting from misperceptions, predispositions, or other subjective, extraneous influence. Evaluators are usually unaware that they are making these judgment errors, but once they become familiar with the common obstacles, they can usually correct themselves and achieve a functional level of accuracy in rating performance.

The following paragraphs contain descriptions and examples of the most common types of bias which can adversely affect the rating process. As you read these paragraphs, try to think of other examples of each kind of judgment error from your own personal experiences. This will make the material more meaningful to you as an individual, and help you understand your personal thinking processes in making judgments of others.

Comparison effects is an error in judgment which results from the tendency for a rater to evaluate a person relative to other individuals rather than on the requirements of a job or task. A rating should be given on the basis of the criteria established prior to the rating process and specified on the rating form, **not** on the basis of a comparison with others.

For example, think of the best looking man or woman you have ever met. Rate this individual on a 7-point scale, with 7 representing outstanding in terms of physical attractiveness. Now think of your favorite glamorous movie star. Rate the movie star on the same 7-point scale you used to rate the previous individual. Now re-rate the first person. If you just lowered the rating of the first person, you have made a comparison error. Each person should be rated on the degree to which they fulfill your predetermined criteria for physical attractiveness, not on how he or she compares with others.

Comparison error is especially troublesome in performance evaluation conducted for the purpose of training enhancement. The team or individual you will be evaluating need to know how they perform in relation to a specific criteria, not how they compare to other teams or individuals. You may have observed someone in the past who performs a particular task extremely well, or extremely poorly. If you use this past experience to make comparisons with the individual presently being evaluated, you would be making a comparison error in judgment. If the present

individual achieves an average level of performance on the task, rating him in comparison to the outstanding person would make his performance look bad and result in a lower rating than actually deserved. Conversely, contrasting him with the poorly performing individual would make average performance look superior in comparison, resulting in a higher rating than appropriate. Either way, the evaluation would result in misleading performance information upon which to base plans concerning future training needs.

First-impression error results from the tendency for a rater to make an initial favorable or unfavorable judgment about the person to be rated, and then unconsciously ignore or discard subsequent information, so that the initial impression is supported.

Everyone has a tendency to occasionally form first impressions when initially meeting someone new. For example, if a person you have just met seems reluctant to look at you while speaking or avoids eye contact, you may have a first impression that this is a dishonest person or perhaps has something to hide. During subsequent contact with this person you might erroneously view other behaviors in accordance with your concepts of dishonesty, thus confirming your original conclusion. Perhaps the person is actually scrupulously honest, but failed to look at you or make eye contact because of shyness rather than an untrustworthy nature.

The team or individuals which you evaluate may have trouble "getting off the ground" initially and then go on to better performance, or get off to a flying start only to deteriorate as the exercise proceeds. If you allow these first impressions to color the way you see the remainder of their task performance, the evaluation will be inaccurate.

"Middle of the road" error is committed by the rater who wants to play it safe. This error refers to evaluators who consistently give ratings at, or close to the midpoint of the scale, even when the performance of the person or team being evaluated clearly warrants a substantially higher or lower rating.

Suppose you were asked to rate different food items on a 5-point scale for taste. If you gave an overcooked burger from a fast food restaurant and a special dish from a 5-star restaurant both a rating of 3, you would probably be committing a "middle of the road" error. This is an extreme example, but serves to illustrate the concept. This is not to say that you should never give average ratings. In fact, the performance of

many of the individuals that you evaluate will be average and should be rated near the midpoint of the scale. However, you should stay alert to performance differences that warrant a higher or lower rating.

Severity and generosity errors are committed by the evaluator who is consistently too hard or too easy in rating the performance of others. This type of judgment error is closely related to the "middle of the road" error discussed above. All three errors (middle of the road, severity, and generosity) reflect the failure of the evaluator to recognize and record differences in levels of performance. When everyone is given approximately the same ratings, whether that rating is high, low, or average, the ratings are of no utility in determining training deficiencies and offer no insight as to how well the objectives are being achieved. Rating according to the criteria specified on the rating scale rather than personal, subjective criteria will help you overcome any tendency to make this type of error.

Halo effect refers to the error of making inappropriate generalizations from one aspect of a person's performance to all aspects of the person's performance. For example, if you know that John Smith was a college football quarterback, currently stars on the post rugby team swims, jogs and plays a superb game of golf, you may conceptualize him as a superior all-around athlete. If you later had to rate John's skill as a tennis player, you may be influenced by your other knowledge concerning his athletic ability, and rate him higher than justified. In reality, John may be a poor tennis player. The skills required to play tennis will not actually be related to the skills required to play golf, swim, or play football quarterback even though they all fall under the general category of athletics. This error can also operate in reverse if you allow poor performance on a task to negatively influence your rating of the individual on an apparently similar or related task.

In your duties as evaluator, you may observe that a team or individual performs some category of tasks quite well, or poorly. This should in no way affect your evaluation on future tasks which appear similar or related. Different performance measures are not always as related as they may appear.

To further help you avoid this judgment error, do not listen to comments about the group you are evaluating, or about any of the individuals who comprise the group. Also, if more than one evaluator is making ratings of the training exercise, be certain that all evaluators assign their ratings independently. Group discussion should be avoided until after all observations and ratings are completed.

Similar-to-me-effect is an error which occurs when raters judge more favorably those individuals whom they perceive as similar to themselves. That is, the more closely an individual resembles the rater in attitudes or background, the stronger the tendency of the rater to judge the individual favorably. This seems to occur because of the human tendency to like or think more highly of others who are perceived to be like us rather than unlike us because it is flattering and reinforcing. This effect may be acceptable and adaptive in social situations, but it is a troublesome source of error when evaluating performance.

APPENDIX B
PROCESS RATING SCALE

Problem-Solving Group # _____

Observer Name _____

<u>Scale</u>	<u>Score</u>
I. Getting Organized	_____
II. Information Sharing and Coordination	_____
III. Decision Making	_____
IV. Professionalism	_____
V. Leadership	_____
Total	_____

I. GETTING ORGANIZED

a. Organizing into teams and assigning roles.

- (1) No Roles or teams were established.
- (2) There was a division of task but no clear assignment of individual responsibility.
- (3) Leader assigned roles arbitrarily without consideration of preferences or qualifications.
- (4) Assignments were based on preference alone OR process of determining qualifications was excessively time consuming.
- (5) Leader assigned roles based on preferences and qualifications.

b. Time Management Plan.

- (1) No attempt made.
- (2) Overall time limit mentioned only.
- (3) Discussed need for time plan, various options considered, no decision reached and clearly stated.
- (4) Developed time plan that was not fully adequate OR process was inefficient and overly time consuming.
- (5) Arrived at clearly-defined, workable plan or schedule that was clearly understood by all members.

c. Problem Definition and Restatement.

- (1) Made no attempt at problem definition.
- (2) Made reference to need only, did not accomplish problem definition.
- (3) Reached an unsatisfactory problem definition.
- (4) Problem definition was only adequate OR process was inefficient and time consuming.
- (5) Efficiently arrived at good problem definition which was beneficial to problem-solving process.

d. Preliminary Information Gathering.

- (1) No information from individual team instructions was presented or requested.
- (2) Information was requested but preliminary information gathering process was unproductive.
- (3) Information was provided from 1 or 2 individual team forms only.
- (4) Sufficient information was provided but process was unduly time consuming with too detailed a level of information provided.
- (5) All teams and leader provided sufficient information to begin problem solving.

_____ **Total Scale Score**

II. INFORMATION SHARING AND COORDINATION

Item is scored if information or functional equivalent is received prior to selection of COA.

Information Received By	Original Source
Training Team	
--- Number of devices	Budget
---- Number of trainees and session	Budget
---- Point cutoff scores for units	Leader
--- Device delivery dates	Budget
--- Operator availability dates	Personnel
---- Ratio of weekday to weekend personnel	Personnel
--- Importance of corps training	Leader
---- Device maintenance schedule	Budget
---- Relative importance of active and reserve	Leader
Personnel	
--- Number of devices	Budget
---- Number of operators and TM session	Budget
---- Ratio of AC to RC training planned	Training
--- Device delivery dates	Budget
--- Device maintenance schedule	Budget
--- Importance of AC and RC	Leader
--- Prototype devices available	Budget
Budget Team	
--- Estimate of TDY costs	Training
---- Estimate amount of training required (number of persons units or session)	Training
--- Estimate of personnel costs	Personnel
Leader	
---- List of units not scheduled	Training
_____ Total Scale Score	

III. Decision Making

a. Develop Courses of Action.

- (1) Group did not arrive at any overall COA but worked on completely unintegrated subsolutions.
- (2) Proposed only 1 COA which was accepted with insufficient analysis.
- (3) COA generation proceeded sequentially. A single COA was analyzed for feasibility before alternatives were considered.
- (4) Proposed 2 or more COA that were not sufficiently distinct or were inadequate in concept or group was inefficient and used excessive time in COA generation.
- (5) Proposed 2 or more adequate and distinct COA.

b. Generate Criteria for Selection.

- (1) No criteria established. Group seemed unaware of need for criteria.
- (2) Discussed need for criteria but none established.
- (3) Criteria were proposed but group failed to reach consensus or achieve a common understanding of the criteria.
- (4) At least one relevant criterion was established but clearly relevant criteria were ignored OR process was inefficient and overly time consuming.
- (5) Criteria were relevant, clearly defined and understood, and no clearly relevant criteria were ignored.

c. Analyze Courses of Action and Prepare Estimates.

- (1) No estimates made or estimates based on meager analysis.
- (2) All teams began work but none completed a satisfactory analysis.
- (3) Only 1 team analyzed COAs sufficiently.
- (4) Two (2) teams analyzed COAs sufficiently.
- (5) Each COA was analyzed completely by all teams. Costs and benefits were accurately identified.

d. Selection of Course of Action.

- (1) No decision reached.
- (2) Arbitrary decision reached without consideration of pros and cons.
- (3) Discussed pros and cons of COAs, but decision reached without use of comparative method.
- (4) Used comparative method without establishing relative weights of criteria.
- (5) Used comparative method to select a COA based on an analysis of established, weighted criteria.

_____ **Total Scale Score**

IV. Professionalism

a. Dedication to Task.

- (1) Excessive amounts of irrelevant behaviors or idleness seriously impaired group performance.
- (2) Moderate amounts of irrelevant behavior or idleness had some negative impact on performance.
- (3) Some members failed to maintain attention to task despite prodding.
- (4) All team members worked continuously but some required prodding by others.
- (5) Entire group showed enthusiasm and attention to task throughout.

b. Error Checking.

- (1) No discernable checking of facts, conclusions, methods, etc., occurred.
- (2) Too infrequent error checking clearly resulted in negative effects on solution.
- (3) Error checking was evident but the overall process was inadequate and at least one major error was uncorrected.
- (4) Error checking process was inefficient and overly time consuming.
- (5) Frequent error checking led to positive effect on outcome, without becoming unduly time consuming.

c. Incremental Improvement of Complete Solution.

- (1) Group quit immediately after solution reached.
- (2) Only 1 or 2 members continue to work after solution reached.
- (3) Group continued to discuss final solution without making improvements.
- (4) Group continued working with minor improvements.
- (5) Group continued work after complete solution was reached with significant improvements.

d. Ability of Group to Work Within Deadlines and Manage Time.

- (1) Group failed to establish or meet deadlines or showed no awareness of deadlines.
- (2) Group showed awareness of deadlines but seldom succeeded in meeting deadlines.
- (3) Sometimes met deadlines.
- (4) Usually met deadlines.
- (5) Members performed tasks to meet all established deadlines and schedules.

_____ **Total Scale Score**

V. Leadership

a. Provided Focus and Direction to Effort.

- (1) Group was directionless. Never seemed to know what they should be doing or how to do it. Leader made no attempt to direct or focus effort.
- (2) Leader made suggestions only or gave inadequate guidance. Group was essentially directionless.
- (3) Leader provided some focus, group generally understood what they must accomplish, but not how to do it.
- (4) Leader provided focus and direction generally, but occasionally failed to redirect effort when group was confused or followed inappropriate or non-productive paths.
- (5) Leader always maintained clear focus and direction of group. Quickly brought group on target.

b. Made and Stated Decisions Clearly.

- (1) Appeared unable to make decisions.
- (2) Decisions were not communicated to group or were made reluctantly.
- (3) Attempted to communicate decisions, but left misunderstandings and confusion.
- (4) Most conflicts were resolved but required excessive time.
- (5) Decisions were made appropriately and clearly communicated to group.

c. Conflict Resolution.

- (1) Leader made no attempt to resolve conflicts where leader intervention was clearly indicated.
- (2) Attempts to resolve conflicts failed.
- (3) Conflicts were resolved arbitrarily to detriment of group performance.
- (4) Most conflicts were resolved but required excessive time.
- (5) All conflicts resolved smoothly so that there were no detrimental effects on group performance.

d. Kept Group to Appropriate Time Schedule.

- (1) Leader made no attempt to keep group to any schedule when group clearly required such guidance.
- (2) Leader tried to maintain orderly progress but failed.
- (3) Leader allowed group to miss deadlines and occasionally fall behind so that some tasks were hurried.
- (4) Leader supervised progression through tasks adequately.
- (5) Strong leadership kept group to timely and orderly sequence of tasks. Group had sufficient time for all phases.

_____ **Total Scale Score**

APPENDIX C

INTERRATER RELIABILITY PROGRAM

PROGRAM FOR INTERRATER RELIABILITY USING FORT LEAVENWORTH UNIVAC SYSTEM

```
@RUN/PRT _A_ _B_ / _A_ _B_ , 50,200
@ASG,A _B_ * _C_ .
@ASG,A _D_ .
@BRKPT PRINT$, _D_
@ASG,A SPSSX*SPSSX.
@SPSSX*SPSSX.SPSS-X 30000
SET WIDTH=80
FILE HANDLE _C_ /NAME=_C_ .
DATA LIST FILE= _C_ /GROUP 1-2, ORGAN1 TO ORGAN3 4-9, INF1 TO INF3
12-17, DECIS1 TO DECIS3 20-25, PROF1 TO PROF3 28-33, LEAD1 TO
LEAD3 35-41
RELIABILITY VARIABLES=ORGAN1 TO LEAD3/
SCALE (GETORG)=ORGAN1 TO ORGAN3/
SCALE (INFOR)=INF1 TO INF3/
SCALE (DECIDE)=DECIS1 TO DECIS3/
SCALE (PROFESS)=PROF1 TO PROF3/
SCALE (LEADER)=LEAD1 TO LEAD3
FINISH
@BRKPT PRINT$
EOF
```

Running an SPSSX program requires that you create three files: (1) a data file to hold the raw data (ratings) - the name of this file should be placed in the blank spaces marked "C" in the above program; (2) an SPSSX program file containing the SPSSX Job Control Language and the SPSSX program above; and (3) an empty file to hold the output of the data analysis - the name of this file should be entered in the blank spaces labeled "D" on the above program. In spaces labeled "A" on the above program, you should place your organization's DPFO account number. In spaces labeled "B" you should place your organization's DPFO account name.

Files are created using the following command:

```
@asg,u [Filename].
```

After the file has been created and data, program, etc., entered, the file can be

accessed with the following command:

@ed,u [filename].

SETTING UP THE DATA FILE

The data set should be entered into your file in the following format which corresponds to the data list statement of the SPSSX program.

COLUMN	1-2	4-9	12-17	20-25	28-33	36-41
01	090608	121112	101212	181516	141414	
02	121415	100908	060506	101008	090709	

NOTE: The numbers are notional for purposes of example.

In the columns 1-2, place the group number or other identifier. Columns 4-9 represent the three raters' ratings for the getting organized scale: Rater #1 scored group #1 with 09 points and group #2 with 12 points, Rater #2 rated group #1 with 06 points and group #2 with 14 points, rater #3 rated group #1 08 points on this scale and group #2 15 points. In columns 12-17, the three ratings are entered for the information gathering scale, rater #1 rated group #1 with 12 points and group #2 with 10 points, rater #2 rated group #1 with 11 points and group #2 with 09 points, rater #3 rated group #1 with 12 points and group #2 with 08 points. Columns 20-25 represent the ratings for the decision making scale, columns 28-33 represent the professionalism scale ratings, and columns 36-41 represent the leadership scale ratings.